SUSMAN GODFREY L.L.P.

April 3, 2025

Hon. Ona T. Wang
Southern District of New York
500 Pearl Street
New York, New York 10007

           Re: *The New York Times Company v. Microsoft Corporation, et al.*,
                    No.: 23-cv-11195-SHS: Opposition to OpenAI's Motion to Compel, ECF
                    495

Dear Magistrate Judge Wang:

Plaintiff The New York Times Company ("The Times") opposes OpenAI's request to expand the date range for The Times's Google Mail, Google Drive, and Slack collection. *See* Dkt. 495.

1. **A Blanket Collection from 2018 is Disproportionate to the Needs of this Case.**

OpenAI seeks to expand the date range for The Times's Google Mail, Google Drive, and Slack collection to 2018 for six custodians: (1) A.G. Sulzberger; (2) Meredith Levien; (3) William Bardeen; (4) Greg Miller; (5) Alex Hardiman; and (6) David Perpich. However, this request completely disregards the parties' agreed upon search protocol. The parties first negotiated search terms, custodians, and date ranges for collection in July 2024. As part of those negotiations, The Times represented that it would collect documents from its custodians beginning from January 1, 2022 (ChatGPT was publicly released in November 2022), and Defendants did not object to that proposal. Consistent with the parties' agreement, The Times then collected documents from its custodians from the start of 2022 through May 1, 2024. Subsequently, months later, OpenAI suddenly sought to reopen negotiations regarding the timeframe for collection. It demanded that The Times go back and re-collect documents from five custodians, and The Times did so. OpenAI's yet further demand that The Times once again conduct a collection of an additional four years for six of its most senior custodians' documents and Slack messages is unjustified.

OpenAI claims that it is entitled to documents relating to its statute of limitations of defense—viz., documents that demonstrate The Times's knowledge of OpenAI's unlawful conduct prior to 2022—and "documents related to the impact of Defendants' LLMs on The Times's business model." Dkt. 495 at 1–2. However, OpenAI has not provided any indication of the kinds of documents that it is seeking or why The Times's current production is insufficient to satisfy its needs. Indeed, The Times has already agreed to produce documents going back to 2018 for several of its high-ranking and technology-focused employees: Gordon Diggs, Sam Felix, Chris Wiggins, and Rebecca Grossman-Cohen. It has also agreed to expand its custodial collection for Hannah Poferl back to 2019. Accordingly, OpenAI's request would mean that The Times is required to extend its collection date range to years before the launch of ChatGPT for 11 of its 22 custodians.

Furthermore, to avoid the significant burden of sifting through millions of emails, The Times offered to work with the six requested custodians to directly run simple searches in their

emails for terms like "OpenAI" and "ChatGPT" for the period prior to 2022, terms that would capture documents relating to OpenAI's stated rationale (Google's email system permits these types of searches, but not the complex terms in the negotiated term sheet). However, OpenAI rejected that offer without any explanation.

OpenAI's request to extend collection four years prior to the public release of ChatGPT makes no sense. Specifically, with respect to evidence of impact on The Times, *OpenAI* has documents relevant to its misconduct from well before 2022, when it used millions of The Times's works to develop its products. However, because much of this unlawful activity came before the public release of OpenAI's GPT product in 2022, The Times is unlikely to have custodial documents relevant to impact from this earlier period. Moreover, OpenAI's assertion that it is simply trying to gather "the facts from the high-level custodians whose knowledge is most likely to be imputed to the company" is irrelevant. Dkt. 495. The appropriate question is which individuals are likely to have responsive documents, and OpenAI has not provided any basis for believing that such high-level executives would have been discussing ChatGPT before its public release. Accordingly, OpenAI's insistence that The Times collect and cull through millions of emails and Slack messages in the hopes that something relevant exists is entirely unreasonable.

Finally, The Times is already producing non-custodial documents related to the harm it incurred—documents related to revenue, subscriptions, and web traffic—going back to 2018 and, in some instances, 2015. In other words, The Times is already producing "responsive documents related to the impact of Defendants' LLMs on The Times's business model." *Id.* Thus, given the extensive scope of The Times's production, expanding the date range for The Times's Google Workspace and Slack collection is not proportional to the needs of this case. *See* Fed. R. Civ. P. 26(b)(1).

2. <u>OpenAI Fails to Explain Why it Cannot Offer Targeted Search Terms for Google Mail Messages.</u>

In its March 28, 2025 email to OpenAI, The Times indicated that from 2018 to 2021, "more than a million messages [were] stored in Google Mail" for these six custodians. Ex. 1 (3/28/25 Muttalib email). In an effort to alleviate the extensive burden of pulling and reviewing these emails, The Times requested that OpenAI send "a narrow set of targeted search terms that [OpenAI] wanted [The Times] to run against this data for [its] expanded date range," which could be run without conducting a full document collection. *Id.* Rather than engaging with this offer, OpenAI stated that The Times's email did not "make clear that the million messages' stored in Google Mail belong[ed] to the six custodians at issue"—an issue that The Times could have explained, had OpenAI simply asked. Moreover, instead of providing search terms, OpenAI indicated that it would "be raising the issue with [the Court,]" and chastised The Times for not providing "hit counts," despite not sending any terms for The Times to run in the Google environment. *Id.* (3/31/25 Salomon email).

The Times can only properly substantiate its burden claim by evaluating OpenAI's proposed search terms and the hit counts associated with their requests. And, OpenAI's assertion that all negotiated search "terms should be applied to [this] custodial collection" is a nonstarter, including because running these complex terms would require the expensive process of pulling

each custodian's files in full and paying the associated costs to host those documents. Dkt. 495; *see also* Dkt. 509. Indeed, OpenAI itself has indicated that this expanded collection is limited to establishing its statute of limitations defense. As such, The Times cannot understand how it is proportionate to run all negotiated search terms for OpenAI's 300 RFPs to achieve this limited end.

In an effort to resolve this issue, The Times ran a broad search across the relevant data for documents relating to ChatGPT and OpenAI, searching for ChatGPT, ChatGPT*, GPT*, GPT, OAI*, OAI, "Open AI," OpenAI, and OpenAI*. None of the documents were responsive, including for OpenAI's statute of limitations defense or to the impact of its LLM products on The Times. Accordingly, the Court should deny OpenAI's request to expand the date range for Google Mail messages. Alternatively, to the extent that the Court believes that an expansion of the date range for Google Mail messages is warranted, it should order OpenAI to offer whatever additional narrow, simple search terms it thinks directly bear on its statute of limitations defense.

3. OpenAI's Request for Google Drive and Slack Data is Disproportionate.

OpenAI claims that it "has repeatedly asked The Times to explain the [alleged] discrepancy between its representations in court regarding burden and its Rule 30(b)(6) custodian's testimony." Dkt. 495. Specifically, OpenAI asserts that "The Times should be able to run search terms in the 'native environment,' given that Google Vault explicitly allows for that capability." *Id.* However, as The Times has repeatedly explained, it cannot run OpenAI's complex terms and obtain hit counts, absent collection, across emails (which can only accommodate simple search strings such as the one above) and especially *Drive data* in Google Workspace (which cannot be searched in this manner at all). Indeed, OpenAI impliedly acknowledges this by noting that Google Vault only allows for the searching of emails by "keyword." Dkt. 495 at 3.

Similarly, The Times's ability to generate hit counts for Slack data without transferring the custodian's full files to a document vendor is limited, as searches time out where there is a significant volume of data. Indeed, despite OpenAI's suggestion to the contrary, The Times cannot use Google Vault to run search terms across Slack data—as Slack messages are not housed in Google Workspace. Moreover, OpenAI's assertion that The Times can simply ameliorate the multiple issues that Slack collection has posed with a date overlay is unfounded. The Times has spent significant time and expense fixing the multiple threading issues that Slack collection has posed for existing custodians based on the parties' longstanding agreement on search terms and date ranges, and it is unreasonable for OpenAI to demand at this late date that The Times bear further expense for this proposed fishing expedition. Accordingly, to the extent the Court is inclined to grant any part of OpenAI's request, The Times requests that the Court limit the expansion of the date range to Google Mail messages.

Respectfully submitted,

/s/ Ian B. Crosby
Ian B. Crosby
Susman Godfrey L.L.P.

/s/ Steven Lieberman
Steven Lieberman
Rothwell, Figg, Ernst & Manbeck

cc: All Counsel of Record (via ECF)